Supporting Online Material

"Adaptive Radiation from Resource Competition in Digital Organisms"

S. S. Chow, C. O. Wilke, C. Ofria, R. E. Lenski, C. Adami

Materials and Methods

Software: All experiments were performed using version 1.99 of the Avida software (*S1*), running under the Linux operating system on a Beowulf cluster of 64 Pentium III processors. Avida is freely available from http://www.dllab.caltech.edu/avida. The configuration files necessary to reproduce these experiments are available at http://www.dllab.caltech.edu/pubs/science04.

Logic functions: The nine resources corresponded to the following one- and two-input logic operations: NOT, NAND, AND, OR_N, OR, AND_N, NOR, XOR, and EQU (see Table 1 in Ref. *S2*). An organism successfully performs a logic function if it reads bit-strings from its environment, performs the computation, and returns the correct result. The Avida software checks the returned values and bestows a commensurate amount of CPU time on the successful organism. The amount of CPU time depends on the function performed and on the availability of the associated resource. If a resource has been depleted, the reward is reduced for the associated computation. Resources are represented by arbitrary real-valued units. An organism can obtain a maximum of 5 units or 0.25% of the total concentration, whichever is smaller, per completed computation. Flow rates of resources are expressed as units per update.

Adaptive radiation of unevolved ancestors: We inoculated populations of size N=3000 with a hand-written ancestor that could self-replicate but not perform any logic functions. Population size was held constant in all experiments. Each experiment ran for 4×10^5 updates (an arbitrary unit of time during which each organism in a population executes, on average, 30 instructions), which allowed between 5,000 and 70,000 generations depending on the functional complexity of the organisms that evolved. For the experiments summarized in Fig. 1, we ran 25 replicates at each of the seven inflow rates; at two inflow rates (0.1 and 10), all the replicates ran for an additional 4×10^5 updates to examine whether species

richness had saturated. In all replicates, unused resources flow out at a rate of one percent of their standing concentration per update. Organisms were allowed to consume a single resource multiple times (if available) throughout their life cycle. As a control, we ran 50 replicates with unlimited resources, where we expect only a single species to exist owing to the absence of any opportunity for frequency-dependent selection; these controls served to calibrate the species-clustering algorithm described below. Organisms in all experiments in this study were constrained to a genome length of 100 instructions by disallowing insertion and deletion mutations. The copy mutation rate was set to 0.005 per copied instruction.

Adaptive radiation of evolved ancestors: For the experiments summarized in Fig. S4, we ran five replicates for each of five different generalist ancestors (25 replicates total) at the same flow rates used with the unevolved ancestor. The five generalist ancestors had the ability to perform all nine logic functions, and had evolved in an environment where resources were unlimited but each resource could be consumed only once in an organism's life cycle. In such an environment, generalists evolve readily (*S2*).

Invasion experiments: For the experiments summarized in Figs. 2 and S1, we took the most abundant genotype from each species shown in Fig. 3 as representative for that species, and carried out six sets of invasion experiments (5 replicates each with different random number seeds). For each set of experiments, we gave one species an initial abundance of 5 organisms, and the remaining species each an initial abundance of 599 organisms (for a total of 3000 organisms). The inflow rate was 10^3 resource units per update. In all cases, the results from the 5 replicates were essentially identical. For the experiments summarized in Fig. S5, we competed these six species representatives with the representative of a species from a single-species ecosystem evolved at inflow 10^5. We carried out competitions at inflow rates of 10^3 and 10^5. At inflow 10^3, we gave the former six representatives each an initial abundance of 50 and the latter one an initial abundance of 2700. At inflow 10^5, we gave the former representatives an initial abundance of 490, and the latter one an initial abundance of 60. The mutation rate was set to zero in all invasion experiments.

Population size: For the experiments summarized in Fig. S6, we carried out experiments identical to those described under "Adaptive radiation of unevolved ancestors" (including the necessary calibration runs) with population sizes N = 30, 100, 300, 1000, and 3000 (the last corresponding to the same experiments as in the main text). For each N, we chose the clustering cutoff c such that the probability of finding $\Delta s_2 \geq c$ was $\leq 25\%$ (see next section for the definition of c and Δs_2). We scaled the range of inflow rates with population sizes such that, for example, an inflow rate of 10^4 resource units per update at population size 3000 corresponds to an inflow rate of 10^3 at population size 300 (that is, in these two cases the amount of resource flowing in per organism is the same). Stated differently, in the experiments shown in Fig. S6, total population size is made proportional to total resource inflow, subject to a conversion factor (which is varied) that specifies the number of organisms supported per unit of resource inflow.

Clustering algorithm: We used a clustering algorithm to separate organisms into species. First, we sorted all organisms in a population such that the successive sums $s_i = \sum_{k=i+1}^{N} \min_{j \leq i} d(g_j, g_k)$ were minimized; g_i denotes the genotype of the ith organism after sorting, and $d(g_i, g_j)$ is a suitable distance function. The organisms i with $\Delta s_i = s_{i-1} - s_i \geq c$, where c is the cutoff, served as barycenters of clusters and thus defined separate species. (We assumed that $\Delta s_1 \geq c$ always.) We grouped organisms j for which $\Delta s_j \leq c$ with the species of the respective organisms i that minimized $d(g_i, g_j)$. For the distance function $d(g_i, g_j)$, we used phylogenetic distance, defined as the total number of different intermediate states between two organisms along the lines of descent leading to their most recent common ancestor. Alternative choices were unsatisfactory; Hamming distance tended to lump organisms that had important phenotypic differences into the same species, and time-to-most-recent-common-ancestor sometimes split organisms with identical phenotypes into different species. We used the 50 control experiments with unlimited resources, where the absence of frequency-dependent selection precludes the evolution of stably coexisting species, as a baseline for determining the cutoff value, by measuring the distribution of

Δs_2, Δs_3, and Δs_4 at 400,000 updates. We chose a cutoff value of c = 151,467, for which the probability of finding $\Delta s_i \geq c$ was $\leq 25\%$ for Δs_2, $\leq 2.26\%$ for Δs_3, and $\leq 0.1\%$ for Δs_4 in these calibration runs. Thus, we had a less than 25% chance of finding two species in the calibration runs, and a less than 2.5% chance of finding three or more species. (The probabilities of finding $\Delta s_i \geq c$ at earlier time points were approximately the same or smaller.) Figure S7 shows that the particular choice for the cutoff does not affect the overall relationship between inflow rate and species richness.

Supporting text

To test the dependence of our results on the properties of the ancestral organism, we have performed experiments identical to our earlier ones, except using as the ancestor in each additional experiment one of five generalist species that can perform all nine logic functions. Each generalist has been obtained in a separate experiment (see Materials and Methods). Although these generalists can perform all logic functions, they are not adapted to intermediate resource availability, having evolved in an environment with unlimited resource availability. Figure S4 shows the same association between maximum diversity at intermediate productivity that we observe using the simple ancestor unable to perform any logic functions, although the peak associated with the generalist ancestors occurs at somewhat higher resource inflows than the peak obtained using the unevolved ancestor (Fig. 1).

We have also tested the stability of communities under invasion, by competing communities evolved at different productivity levels. Our general result is that the number of surviving species in these experiments is determined by the productivity level at which the competition is carried out (Fig. S5), but that the particular species evolved at the given productivity level does not necessarily survive. For example, in the right panel of Fig. S5, the ecosystem evolved at intermediate productivity contains one species that is a better competitor at high productivity than the native species.

The Avida software keeps the total population size N constant at all times; therefore, we cannot directly test how species richness would respond if total population size varied dynamically—evolutionarily as

well as ecologically—as a function of resource inflow rates. However, we have carried out additional experiments at various fixed N, in order to assess the influence of population size per se. In these experiments, we scaled the range of resource inflows that we used with N (see Methods). Fig. S6A shows that the overall relationship between inflow rate and species richness does not depend on population size, as long as the population size is not too small. Very small population sizes (on the order of a hundred organisms or fewer) cannot sustain more than one or two species. When we plot species richness as a function of population size for intermediate resource inflow rates per organism (Fig. S6B), we see a clear trend towards increased species richness with increasing population size. However, it is not clear whether this trend will continue beyond $N = 3000$, or whether species richness will level off or even start to decrease again. Limitations in computational resources preclude us from running substantially larger population sizes at this time.

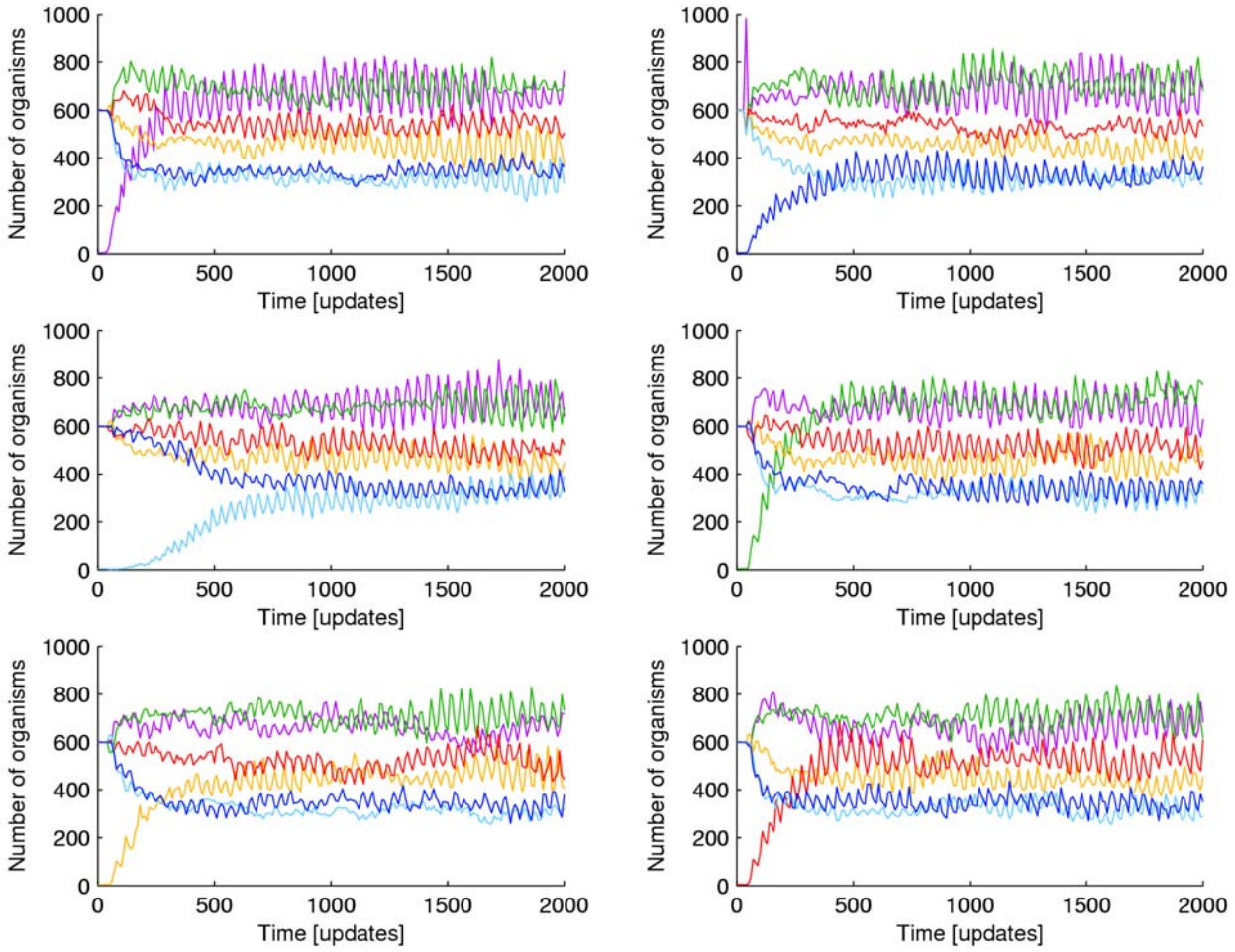

Figure S1: Invasion when rare with numbers of organisms shown as a function of time. Colors correspond to the same six species as in Figs. 2 and 3. Each of the six species in this example ecosystem can invade when rare, and all species quickly settle into an equilibrium abundance.

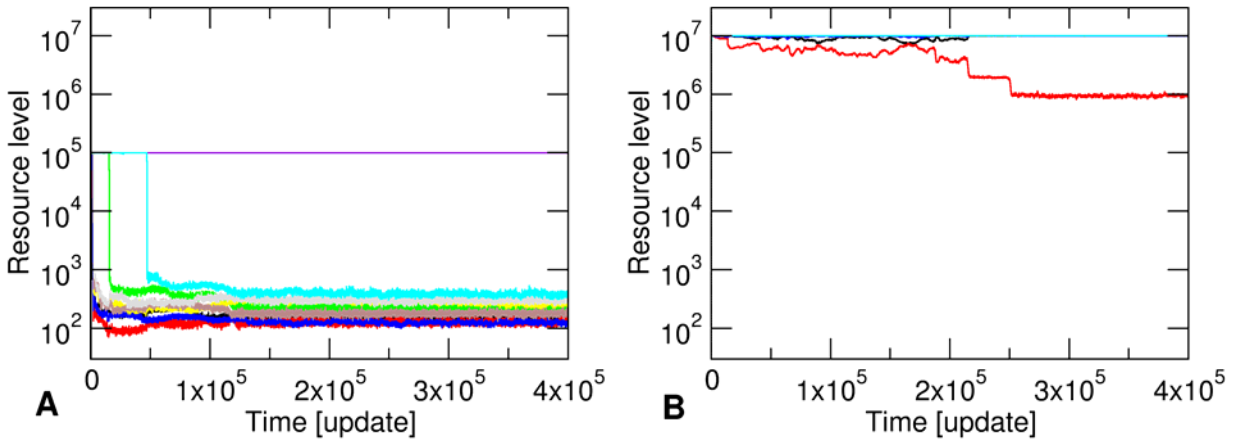

Figure S2: Resource abundance as a function of time. **A**: Typical resource usage at an inflow rate of 10^3 resource units per update for the same exemplary ecosystem from which Figs. 2-4 in the main text were generated. All resources except one are eventually drawn down to a level below the 10^3 units that flow into the system in every update. This low equilibrium abundance indicates that resources are scarce, and that no species can substantially increase the amount of any resource it is using. **B**: Typical resource usage at an inflow rate of 10^5 resource units per update. After a transient period during which several resources are used, a single-species community settles down to using only a single resource (the resource associated with the function NAND). This resource remains well above the resource inflow per update, such that all organisms have an ample supply of resources.

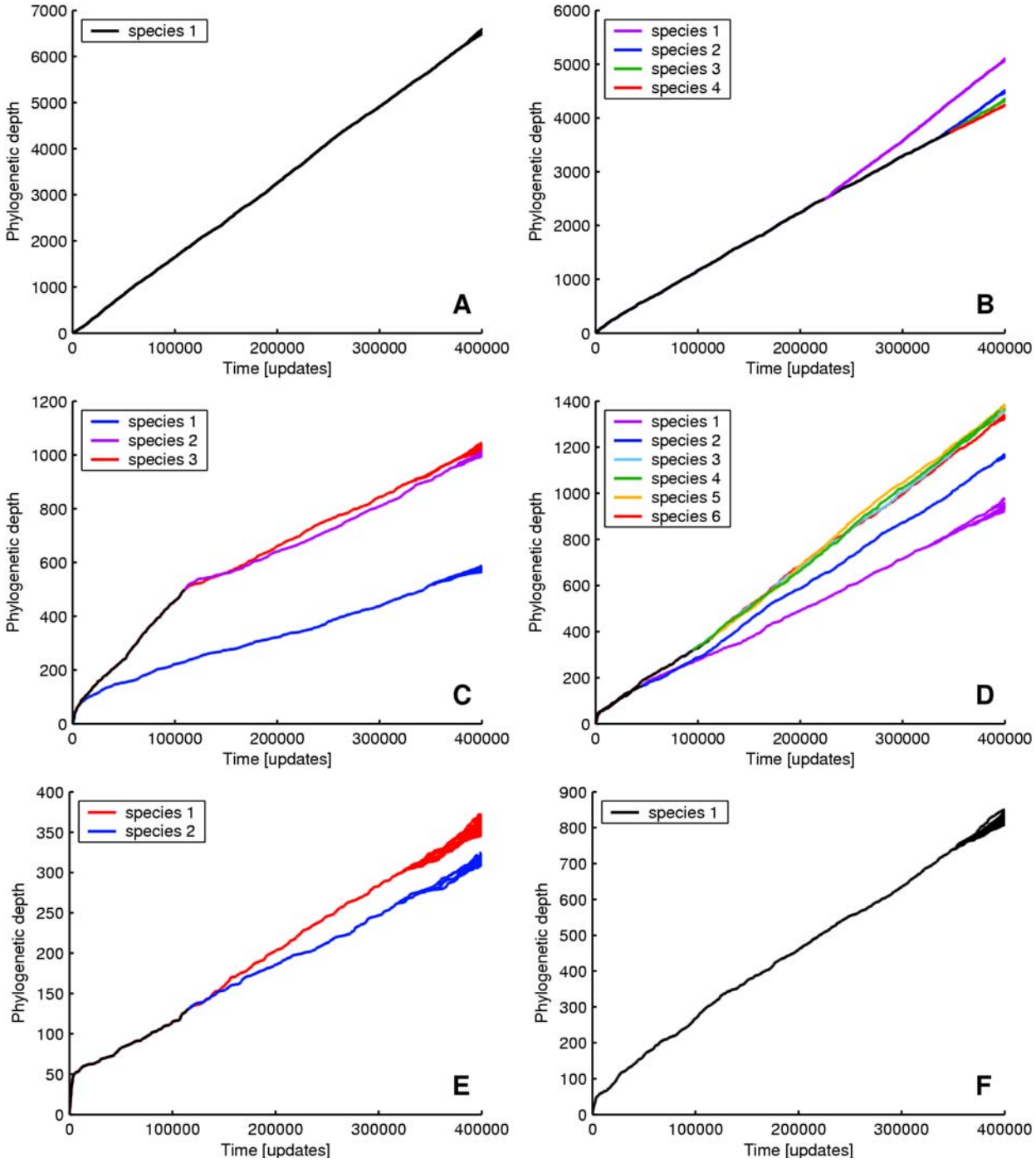

Figure S3: Typical phylogenetic depth profiles as a function of time for various productivity levels. Inflow rates are: 10^0, 10^1, 10^2, 10^3, 10^4, 10^5 (A-F respectively). Only phylogenies from experiments at intermediate productivity levels show deep, coexisting branches that indicate speciation.

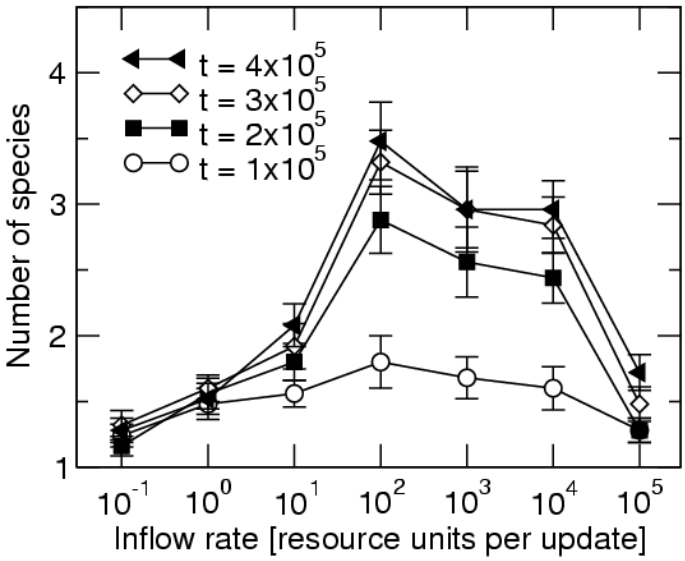

Figure S4: Mean number of species as a function of the inflow rate, starting from a generalist ancestor that could use all of the different resources. Each point is based on 25 runs, with five replicate runs starting from each of five independently evolved generalists. Bars indicate standard errors.

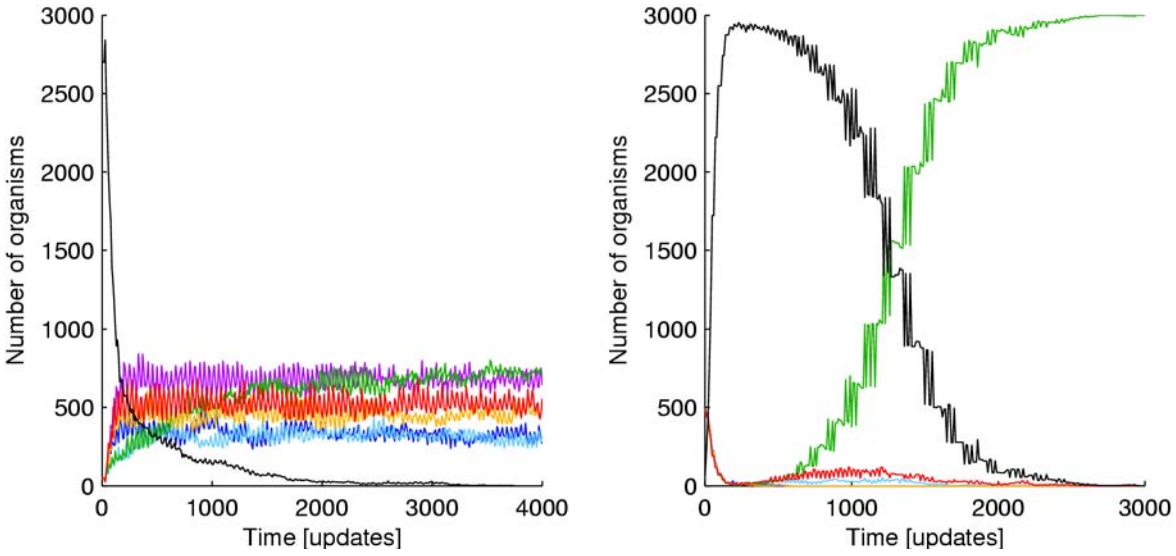

Figure S5: Invasion when rare, with numbers of organisms as a function of time. Left panel: At inflow rate 10^3, the multi-species ecosystem evolved at that inflow rate (colored lines) quickly replaces the single species evolved at inflow rate 10^5 (black line). Right panel: At inflow rate 10^5, the species evolved at that inflow rate (black line) initially suppressed the ecosystem evolved at inflow rate 10^3 (colored lines). However, the ecosystem contains one species (indicated in green) that competes more efficiently at inflow rate 10^5 than the native species, and which eventually replaces the native species.

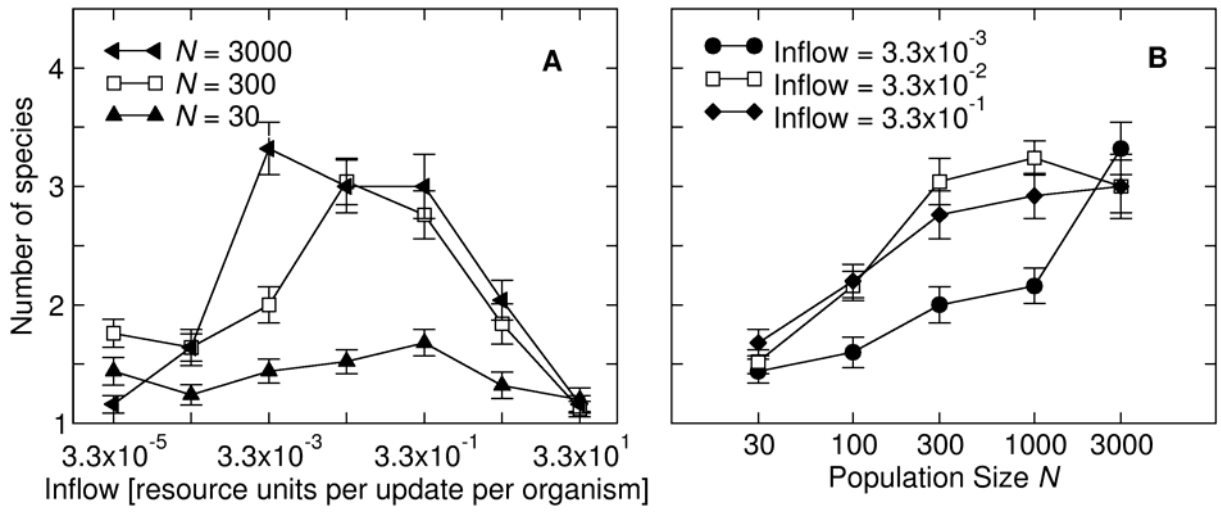

Figure S6: Mean number of species at update 400,000 as a function of inflow rate per organism (**A**) and population size set proportional to total resource inflow (**B**). Each point represents 25 replicate runs, with bars showing standard errors. The total resource inflow rates were scaled with population size *N*, to keep constant the amount of resources potentially available to each organism across different population sizes.

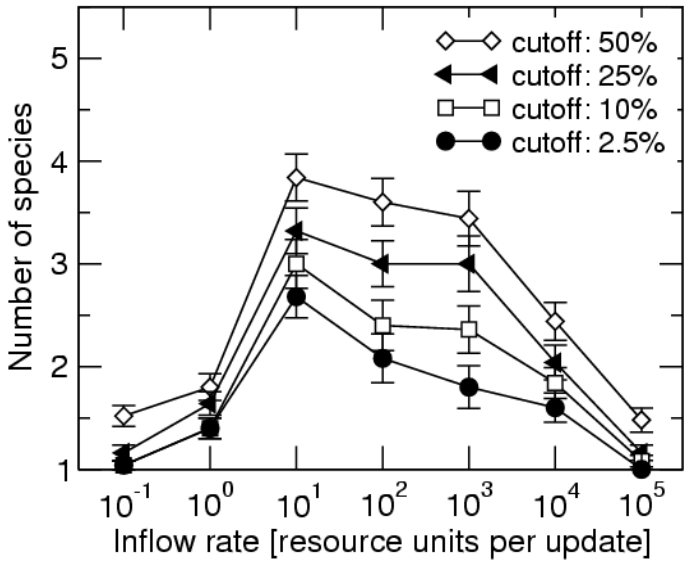

Figure S7: Mean number of species as a function of inflow rate at update 400,000 for various cutoff values. Cutoff percentages indicate the probability of spuriously finding a second species in the infinite-inflow experiments. Each point represents 25 replicate runs, with bars showing standard errors. The data set for cutoff 25% corresponds to the one shown in Fig. 1 in the main text.

References

S1. C. Ofria and C. O. Wilke, *Artificial Life* **10**, 191 (2004).
S2. R. E. Lenski, C. Ofria, R. T. Pennock, C. Adami, *Nature* **423**, 139 (2003).